Bootstrapping Food Preferences Through an Adaptive Visual Interface

*Longqi Yang*, Yin Cui, Fan Zhang, JP Pollak, Serge Belongie, Deborah Estrin
MOTIVATION

Food preferences learning is important!
Health and Life

Obesity  113M
HBP  50M
Diabetes  15M

Unflavored Healthy diet recommendations are of NO Benefit!

*Number of Americans Living with Diet- and Inactivity-Related Diseases
Social Media and Commerce

Personalized diet profile is the Key to user experience!
Our Vision

- Social Network: Content personalization
- Online Groceries: Customers Targeting
- Recipes: Food environment at home
- Restaurants: Customized dishes
- Clinicians: Personalize Treatment Plan
- Nutritionists: Healthy recommendations

Personalized Diet Profile
OUR SOLUTION

An adaptive visual interface
Exploration, 2 iters

Start

Diet Profile

Exploration–exploitation: <15 iters

10 food items

Pairwise Comparison

Take a look at the food below and tap all that you like.

Done

Compare the food pair below and tap on whichever you prefer.

(Press on ‘Like’ if neither of them fits to your taste)
✓ Efficient: completed within a minute.

✓ Visual interface: low cognitive load, personalized and legible.

✓ Preference Elicitation: NO history required, NO ratings.

✓ Deep understanding of food images.

✓ Novel Online Learning Framework.
System Design

Online Learning

- Online Learning framework (LE + EE)
  - What images to present to the user?
  - How to update users’ preferences?

Food Similarity Embedding

- Users have close preferences for similar items
  - Feature representation that can reflect similarities

Food Items Harvesting

- Food images and metadata.
System Design: offline

Food Items Harvesting

- 12,000 food items from Yummly API.
- Images + Metadata (ingredients, nutrients etc.)
- Outliers filtering, 10,028 items were used.
System Design: **offline**

**Food Similarity Embedding**

Representation: 1000 dim visual + 200 dim ingredients

1000 dim visual feature from *Food-CNN*

<table>
<thead>
<tr>
<th>Food Items Harvesting</th>
<th>User Preference</th>
<th>Food already explored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-trained Deep Siamese Network</td>
<td>Ingredients</td>
<td></td>
</tr>
<tr>
<td>Image raw pixels</td>
<td>Metadata</td>
<td></td>
</tr>
</tbody>
</table>

**Contrastive Loss**

$$f(x)$$

$$f(y)$$
System Design: \textit{offline}

\begin{itemize}
  \item \textbf{Offline:}
    - Pre-trained Deep Siamese Network
    - Image raw pixels
    - Metadata
    - Food Items Harvesting

  \item \textbf{Online:}
    - Visual User Interface
    - User Preference
    - Food already explored

  \item \textbf{Backend Online Learning}
    - Exploration (2 iterations)
    - Exploration – Exploitation (>2)

\end{itemize}

\textbf{Food Similarity Embedding}

Representation: **1000 dim visual** + **200 dim ingredients**

\begin{align*}
  &\{ \text{\ding{116}}, \text{\ding{120}} \} \quad l = 1 \quad \| \text{\ding{116}} - \text{\ding{120}} \| \approx 0 \\
  &\{ \text{\ding{113}}, \text{\ding{119}} \} \quad l = 0 \quad \| \text{\ding{113}} - \text{\ding{119}} \| > m
\end{align*}

\begin{align*}
  \mathcal{L} &= \frac{1}{2}lD^2 + \frac{1}{2}(1 - l)\max(0, m - D)^2 \\
  \{ \text{\ding{113}}, \text{\ding{119}} \} \quad l = 0 \quad \| \text{\ding{113}} - \text{\ding{119}} \| > m
\end{align*}

Pairs/Labels were sampled from Food-101 dataset.
System Design: **offline**

**Food Similarity Embedding**

Representation: **1000 dim visual + 200 dim ingredients**

200 dim ingredients feature

- Lemmatization and preprocessing.
- Filtering: Top 200 ingredients.
- Feature vector: 0–1 vector denotes the existence of the ingredient.

Visual and ingredients feature vectors are normalized separately with $l_1$ norm.
System Design: *online*

**Online Learning**

Food preferences representation:

$$p^t = [p_0^t, p_1^t, ..., p_{|\mathcal{S}|}^t] \quad \sum_{i} p_i^t = 1$$

*Distribution of preferences* over all food items in $\mathcal{S}$

$p^t$: updated preference vector after *iteration* $t$

Two tasks at each *iteration* $t$:

- **User state update:** update $p^t$ based on the items presented and user’s choices at *iteration* $t-1$.
- **Images selection:** Select a set of images to show at *iteration* $t$. 
System Design: *online*

- **User state update:**
  
  Update $p^t$ based on the items presented and user’s choices at *iteration $t-1$.*

<table>
<thead>
<tr>
<th>Users’ selections</th>
<th>Image Labeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images selected</td>
<td>Label “+1”</td>
</tr>
<tr>
<td>Images not selected</td>
<td>Label “−1”</td>
</tr>
<tr>
<td>Images not presented</td>
<td>Label “0”</td>
</tr>
</tbody>
</table>
System Design: *online*

**User state update:**

update $p^t$ based on the items presented and user’s choices at *iteration $t-1$*. 

\[
\min_u \sum_{j=1,j \neq i}^{\mathcal{S}} \omega_{ij} (y_i - u_j)^2 + \sum_{j=1,j \neq i}^{\mathcal{S}} (1 - \omega_{ij})(u_j - y_j)^2
\]

**Smoothness**

**Fitting**

*Label Propagation and Exponentiated Gradient Algorithm (LE)*

\[
\omega_{ij} = e^{-\frac{1}{2\alpha^2} \| f^s_i - f^s_j \|^2} \quad u_j = \sum_{i=1}^{\mathcal{S}} \omega_{ij} y_i \quad p_i^t \leftarrow p_i^{t-1} \times e^{\frac{\beta u_i^{t-1}}{p_i^{t-1}}}
\]
# System Design: online

<table>
<thead>
<tr>
<th>Online Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Images selection:</strong> Select a set of images to show at <em>iteration t</em>.</td>
</tr>
<tr>
<td><strong>Exploration and Exploration–exploitation Algorithm (EE)</strong></td>
</tr>
</tbody>
</table>
| **Exploration (Ten images):** $t \leq 2$  
  K-means++ |
| **Exploration–exploitation (Two images):** $t > 2$  
  One Item that user “prefer” (with high value of $p$)  
  The other item that user hasn’t explored. |
System Design: online
EXPERIMENTS AND USER STUDY

Evaluation, findings and evidence
Experiments: embedding

Clustering performance of Food-CNN (Tested on Food-101 dataset).

- $K$-neighbors of each test image, calculate the precision-recall for each $K$.
Experiments: *user study*

- 227 anonymous users.
- Two factors were controlled in the study.

1st. Algorithm:

<table>
<thead>
<tr>
<th>Label Propagation and Exponentiated Gradient (LE)</th>
<th>Exploration and Exploration-exploitation (EE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Perceptron (OP)</td>
<td>Random Selection (RS)</td>
</tr>
</tbody>
</table>

2nd. Number of iterations: 5/10/15
Experiments: *user study*

- Algorithm to test: *LE+EE*
- Trials: *1/3*

One image from top 1% of preference value. (*unexplored*)
The other image from bottom 1% of preference value. (*unexplored*)
Experiments: *user study*

- Algorithm to test: *LE+EE*
- Trials: 2/3

**Exploration**  
**Exploration–exploitation**

- PlateClick (5 iters)  
- Testing (10 iters)

One image from top 1% of preference value. *(unexplored)*  
The other image from bottom 1% of preference value. *(unexplored)*
Experiments: user study

- Algorithm to test: LE+EE
- Trials: 3/3

Exploration Exploration-exploitation

PlateClick (15 iters) Testing (10 iters)

One image from top 1% of preference value. (unexplored)
The other image from bottom 1% of preference value. (unexplored)
Experiments: *user study*

Prediction accuracy under different algorithms and number of iterations
Experiments: user study

Cumulative distribution of prediction accuracy for LE+EE algorithm
Conclusions and Future work

- Engine for food preferences learning.
- Applicable to general human-in-the-loop problems.